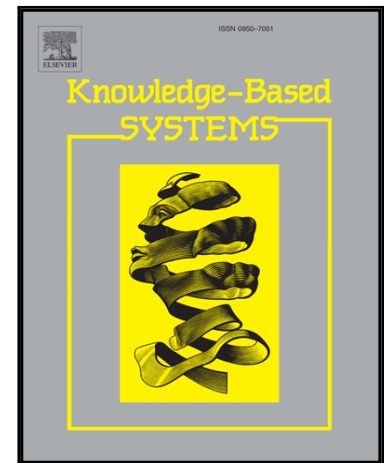


Cross-Document Event Ordering through temporal, lexical and distributional knowledge

Borja Navarro-Colorado, Estela Saquete

PII: S0950-7051(16)30247-7
DOI: [10.1016/j.knosys.2016.07.032](https://doi.org/10.1016/j.knosys.2016.07.032)
Reference: KNOSYS 3620



To appear in: *Knowledge-Based Systems*

Received date: 2 March 2016
Revised date: 20 May 2016
Accepted date: 24 July 2016

Please cite this article as: Borja Navarro-Colorado, Estela Saquete, Cross-Document Event Ordering through temporal, lexical and distributional knowledge, *Knowledge-Based Systems* (2016), doi: [10.1016/j.knosys.2016.07.032](https://doi.org/10.1016/j.knosys.2016.07.032)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Cross-Document Event Ordering through temporal, lexical and distributional knowledge

Borja Navarro-Colorado, Estela Saquete

*Natural Language Processing and Information Systems Group
University of Alicante, Alicante, Spain*

Abstract

In this paper¹ we present a system that automatically builds ordered timelines of events from different written texts in English. The system deals with problems such as automatic event extraction, cross-document temporal relation extraction and cross-document event coreference resolution. Its main characteristic is the application of three different types of knowledge: temporal knowledge, lexical-semantic knowledge and distributional-semantic knowledge, in order to anchor and order the events in the timeline. It has been evaluated within the framework of SemEval 2015. The proposed system improves the current state-of-the-art systems in all measures (up to eight points of F1-score over other systems) and shows a significant advance in the Cross-Document Event Ordering task.

Key words: event ordering, temporal information processing, cross-document temporal relation, cross-document event coreference, timelines, distributional semantics

1. Introduction

The problem of cross-document event ordering consists of extracting events involving a particular target entity among different documents, and ordered them chronologically in a timeline [43].

Email addresses: borja@dlsi.ua.es, stela@dlsi.ua.es (Borja Navarro-Colorado, Estela Saquete)

¹This paper has been partially supported by the Spanish government, project TIN2015-65100-R and project TIN2015-65136-C2-2-R.

As human beings, we tend to organize the flux of happening in structured units known as events. Each event is a fact that occurs in the (real or imaginary) world with a specific structure (the event structure) [26], and denotes processes, activities, states, achievements or accomplishments [38]. An event involves participants [28] and other components that complete the event such as time, place, instruments, patients, etc. (see Example 1). In ISO TimeML Working Group [24], the event was defined as “something that can be said to obtain or hold true, to happen or to occur”.

- (1) <A0>The Airbus A380, the world’s largest passenger plane</A0>, was set to <EVENT eid="e78">land </EVENT> <LOC>in the United States of America</LOC> <TIMEX>on Monday</TIMEX> after a test flight.

The fact expressed in this sentence (1) is a landing, where “to land” (EVENT tag) is the head of the event. The abstract event structure is a flying object (a plane, the Airbus A380 A0 tag) that lands in a place (a country, the United States of America LOC tag) in a specific time (on Monday TIMEX tag). The event heads are usually verbs, as in this example, but they could also be nouns (nominal events). Moreover, relating and ordering the information extracted from *different documents* is an essential task to obtain this knowledge. This cross-document processing improves the traditional single-document extraction and uses information redundancy to its advantage [28] [39].

Cross-Document Event Ordering implies the accomplishment of three sub-tasks. First of all, the *extraction of events* and related entities from texts, because it is necessary to know which events appear in each document, and which entities are related to each one of them. Then *temporal information processing* is required in order to extract the temporal expressions and the temporal relationships established between these events, determining thus which events happen at the same time. Finally, *cross-document event coreference* is needed in order to cluster all the mentions that refer to the same event, regardless of the words used to express them. Example 2 contains two event mentions that refer to the same fact. In a classical temporal information processing system, they are tagged in the text as two different events (*e1* and *e2*).

- (2) Suspected bombs [exploded *event (eid: e1)(class: occurrence)*] outside the U.S. embassies in the Kenyan and Tanzanian capitals [Friday *timex (class:date)(value:1998-08-07)*]. Terrorists provoked the [blast *event (eid: e2)(class: occurrence)*]

In order to properly relate the information, it is necessary to detect when different event mentions are referring to the same fact and to cluster them together, obtaining also all the related information to this specific fact. In our example, the explosion of bombs on “1998-08-07” was provoked by terrorists.

The final aim of combining event extraction and temporal information processing with cross-document event coreference enables us to automatically build ordered timelines of events from written texts. Considering one specific entity as the target entity, all the events related to the target entity are extracted from several documents and arranged in a timeline² (see Figure 1).

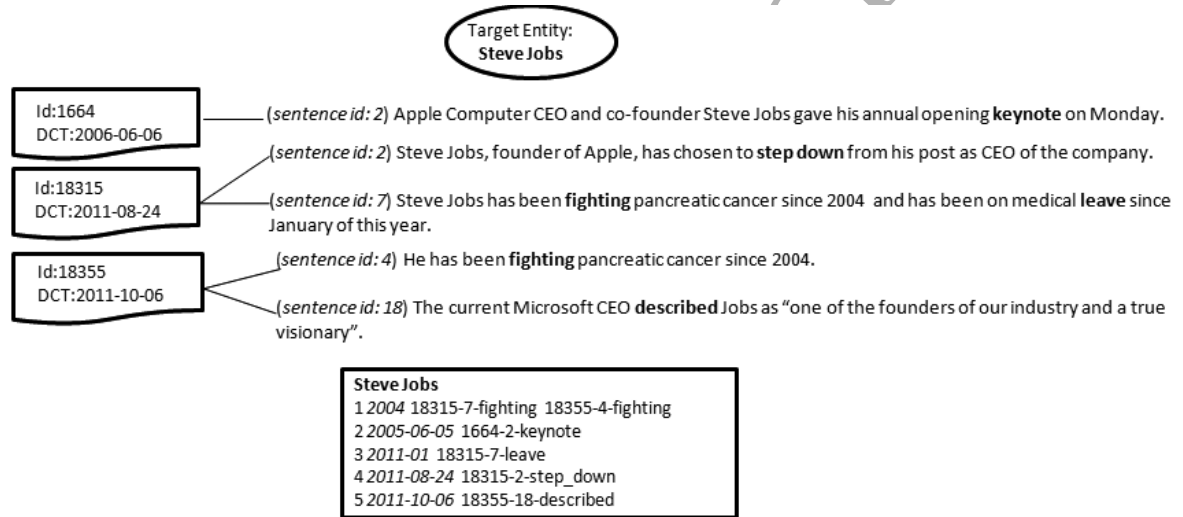


Figure 1: SemEval-2015 Task 4 Example Timeline

In this paper we present a system that copes with this problem and in doing so, our approach is dealing with the three subtasks previously mentioned: a) event extraction, b) cross-document temporal relation extraction and c) cross-document event coreference resolution. Its main characteristic is the application of three different types of knowledge to resolve these problems: temporal knowledge, lexical-semantic knowledge and distributional-semantic knowledge.

Our approach attempts to formalize the idea that two or more event mentions corefer if they have not only temporal compatibility (the events occur

²Example extracted from the SemEval-2015 Task 4 description.

at the same time) but also semantic compatibility (the event mentions refers to the same fact according to their semantic knowledge). In order to detect semantic compatibility, we will analyze two kind of semantics: lexical semantics and distributional semantics. From our point view, both approaches to semantics are compatible: while lexical semantics encode the meaning of words, distributional semantics encode the use of words in specific contexts.

Cross-document event ordering was the topic of the latest SemEval-2015 Task4 [43], called “TimeLine: Cross-Document Event Ordering”.³ In this paper we will use this Timeline task as an evaluation and discussion framework. Compared with the systems presented there, our system achieves the best results and that means a significant advance.

The paper is organized as follow: next Section 2 presents the background on Temporal Information Processing, Event Co-referent Resolution and Cross-Document Event Ordering. Then, in Section 3, we will explain our system and how it resolves the problems previously commented. Section 4 is devoted to the evaluation and discussion of the approach, and finally the main conclusions will be presented in Section 5.

2. Background

2.1. Temporal Information Processing

In general, any task that implies the interpretation of the temporal aspects of language has to handle temporal expressions, events and their relations within the text. Many natural language processing (NLP) areas use temporal information to their advantage. For example, information retrieval (IR), question answering (QA), and text summarization are relevant NLP areas in which the usefulness of specific processing of temporal information has been demonstrated [2] [53] [15]. For these tasks, the automatic processing of the temporal dimension is crucial in order to deal with events that are temporally anchored to some point or period in time.

Several efforts have been made to define standard ways to represent temporal information in texts [23], [18], [56]. At the moment, the most widespread

³This task was partially motivated by the work presented in the European project Newsreader (<http://www.newsreader-project.eu/>). The goal of this project is to reconstruct story lines across news articles in order to provide policy and decision makers with an overview of what happened, to whom, when and where, and timelines are intermediate event representation towards this goal.

one is TimeML schema [52]. It annotates not only events and temporal expressions, but also temporal relations, known as links. Example (3) shows a sentence annotated with TimeML temporal expressions (TIMEX3), events (EVENT), and their links.

```
(3) John <EVENT eid="e1">came</EVENT> on <TIMEX3 tid="t1">Monday</TIMEX3>
    <TLINK eventInstanceID="e1" relatedToTime="t1" relType="IS_INCLUDED">
```

The most recent conferences about temporal information processing and temporal relation extraction were part of the SemEval challenges: *TempEval-1 (2007)*: Temporal Relation Identification [61], *TempEval-2 (2010)*: Evaluating Events, Time Expressions And Temporal Relations [62], and *TempEval-3 (2013)*: Temporal Annotation [60].

All these challenges focused mainly on temporal relations of events, in order to: a) discover which of them occur before, simultaneously or after others, and b) annotate all this temporal information (events, timex and relations) with TimeML annotation scheme. Nowadays there are several Temporal Information Processing systems such as [59], [57], [7], [19], [29], [63], [11], [10], [30] and [37]. An extended version of the last one is included in our system for the event extraction and temporal information processing subtasks.

2.2. Event Coreference Resolution

Events have been a topic of interest in Philosophy [3] [38] and Linguistics [33] [26] from a long time. From a computational point of view, different approaches to automatic event extraction have been proposed, as [28] [1] [41] [37].

There is a recent interest in event coreference resolution, which consists of grouping together the event mentions that refer to the same real-world events into a set of clusters such that all the mentions from the same cluster correspond to a unique event [5].

The first approaches to event coreference for MUC [27] [4] were focused on scenario-specific events. OntoNotes created restricted event coreference [51], corefering only some nominalized events and some verbs. Danlos [16] worked on event coreference at sentence-level. Most of the event coreference approaches are within-document approaches. Chen and Ji [12] proposed to use spectral graph clustering to cluster events. McConky et al. [42] proposed a hybrid approach where the similarity of two events is determined by a combination of the similarity of the two event descriptions, in addition to

the similarity of the event context features of location and time. A dynamic weighting approach was selected to combine the three similarity scores together. Liu et al. [35] presented a supervised approach, describing a method for propagating information between events and their arguments to improve the results.

The cross-document aspect regarding event coreference, however, has not often been explored. Bagga and Baldwin [4] proposed one of the first approaches in this area. Ji et al. [28] worked on a timeline task using the ACE 2005 training corpora. The task was to link pre-defined events involving the same centroid entities (i.e. entities frequently participating in events) on a timeline. Bejan and Harabagiu [6] are performing cross- and within-document approaches and they are using a rich set of linguistic features to model the event structure, including lexical features such as head words and lemmas, class features such as PoS or event class, semantic features such as WordNet sense or semantic roles frames, etc. They use an unsupervised approach based on a non-parametrical Bayesian model. In the work presented by Li et al. [34] the goal was to provide an event fusion approach to obtain the most complete event possible by combining a set of coreference event mentions from different documents which were crawled from Websites. Another cross-document approach is proposed by Lee et al. [32] introducing a novel coreference resolution system that models entities and events jointly. Cybulska and Vossen [13] apply an event model based on four components: location, time, participant and action. They avoid the use of machine learning methods in order to analyze how event components influence event coreference. Goyal et al. [22] use a syntax-based distributional semantic approach on event coreference resolution.

2.3. Cross-Document Event Ordering

Recently, SemEval-2015 [55] included a task that tried to combine the knowledge generated from both tasks previously presented (temporal processing and event coreference) in order to obtain a timeline of events related to a specific given entity, from a set of documents [43]. They proposed two different tracks on the basis of the data used as input. Track A, for which they provided only raw text sources, and Track B, for which they also made gold event mentions available. Each main Track has a Subtrack in order to evaluate only temporal relations between events but not time normalization or event anchoring.

Track A had two participants: **WHUNLP** team and **SPINOZAVU** team. **WHUNLP** team processed the texts with Stanford CoreNLP⁴ [40] and applied a rule-based approach to extract target entities and their predicates. They also performed temporal reasoning.⁵ The **SPINOZAVU** system [9] is based on a pipeline developed in the NewsReader project. It addressed entity resolution, event detection, event-participant linking, coreference resolution, factuality profiling and temporal relation processing, first at document level, and then at cross-document level, in order to obtain timelines.

Track B had also two participants: the **Heideltoul** team and the **GPLSIUA** team. The **Heideltoul** approach [46] uses the HeidelTime tool for temporal information processing and the Stanford CoreNLP for event coreference resolution. Besides, a cosine similarity matching function and a distance measure are used to select which sentences and events are relevant for the target entity. Our own approach, **GPLSIUA** [47], was based on a previous version of the system presented here. It used the OPENER language analysis toolchain for entity detection, the TIPSem tool for temporal processing and a topic modeling algorithm over WikiNews corpus in order to detect event coreference.

Besides, this competition created an interesting evaluation setting in the Cross-Document Event Ordering task. Using this setting, the work presented by Laparra et al. [31] showed that explicit temporal relations are not enough to obtain a full time-anchor annotation of events and evidenced the need for a temporal analysis at document level. In their evaluation, they improved the results presented by **WHUNLP** team and **SPINOZAVU** team at SemEval-2015 Task 4.

The system presented in this paper is an improved version of the one presented at SemEval-2015 Task 4. As we will show later, both systems differ in two main aspects. First of all, the knowledge base used in Semeval version is Wikinews, while the knowledge base used in this version is the complete Wikipedia. Therefore, the version presented at SemEval used a domain-specific knowledge base whereas the system presented here is meant to be a general purpose version. On the other hand, the cluster algorithm used in the SemEval version was a simple K-Mean. It forces the system to define beforehand the number of clusters in which it must group the events

⁴<http://stanfordnlp.github.io/CoreNLP/>

⁵No bibliography is available apart from the general paper of SemEval 2015 Task 4.

together. Due to the fact that it is not possible to know beforehand the number of clusters needed, a different approach has been followed in this version: two event vectors are clustered only if they have semantic similarity. Both aspects will be explained in section 3.1.5.

In the next sections the novelty of our system is presented in depth and evaluated according to the SemEval 2015 Task 4 evaluation framework.

3. Our proposal for Cross-Document Event Ordering

As explained before, given a set of documents and a set of target entities, the task of Cross-Document Event Ordering consists in building an event timeline for each entity. The novelty of our proposal relies on the performance of two clustering methods using different types of knowledge. Each clustering method is able to resolve temporal relation extraction on one hand and event coreference resolution on the other hand. These clustering methods are:

- Temporal clustering: by using the temporal information annotated by a temporal information processing system, the temporal relations between the events are established and the events can be ordered and anchored to the timeline.
- Semantic clustering: the events are clustered using lexical semantics (lemmas and synonyms) and distributional semantic knowledge (topic modeling over Wikipedia) in order to resolve event coreference.

Formally, the main idea of our approach is that two events $e1$ and $e2$ will be coreferent if they have not only temporal compatibility ($e1_t = e2_t$) but also if they refers to the same facts (semantic compatibility: $e1_s \simeq e2_s$):

$$coref(e1, e2) \rightarrow (e1_t = e2_t) \wedge (e1_s \simeq e2_s)$$

Temporal compatibility is obtained from a Temporal Information Processing. About the semantic compatibility between two event mentions, there are two main problems: first, determining which kind of semantic information must be taken into account (lexical semantics, semantic roles, etc.), and second, how this information must be formalized.

Bejan and Harabagiu (2014) [6] proposes a rich semantic characterization of an event mention through a matrix of linguistic features such as lexical features (tokens and lemmas), part-of-speech, TimeML event classes and

Wordnet features (synonyms, lexicographical files) and semantic roles. Other approaches, such as Goyal et al. (2013) [22] proposes a more syntax-based distributional representation of event structure.

Our proposal is focused on two specific semantic aspects of the event structure: the event head on one hand (lexical approach), and the event arguments on the other hand (distributional approach). For each one of these aspects, the system profits from two types of semantic knowledge respectively: lexical knowledge and distributional knowledge.

Distributional semantics is a current event of Computational Semantics which, instead of extracting the sense of a word from a hand-made lexicon or dictionary as in Lexical Semantics, tries to infer word meaning from how the words are used in real texts. The theoretical background of this approach [20, 25] postulates that it is possible to know the meaning of a word by the company it keeps in real contexts (that is, according to the words it tends to appear with in texts). From a computational point of view, this postulate is formalized by representing words as vectors in a multidimensional space. These vectors represent how frequently two words tend to appear together in the same contexts [58].

From a semiotic and cognitive point of view, according to [50, 48], both the lexical and the distributional semantic information needed to understand a text are stored in knowledge bases. Therefore, during the interpretation of a text, there are two cognitive sources of knowledge at work: the lexical knowledge base (or dictionary or Lexicon), that stores the standard or lexical meaning of each word; and the distributional knowledge base, which stores the information about how a word is used [21], that is, to what extent two words usually appear together in the same context. Our approach to event coreferent resolution tries to formalize this cognitive structure: the lexical knowledge is used to find coreferential events according to the event head, and the distributional knowledge is used to find coreferential events according to the arguments of the event structure.

3.1. Architecture of the system

Figure 2 shows the main modules of the system, the external tools used and the knowledge bases. Each module will be explained in the next subsections.

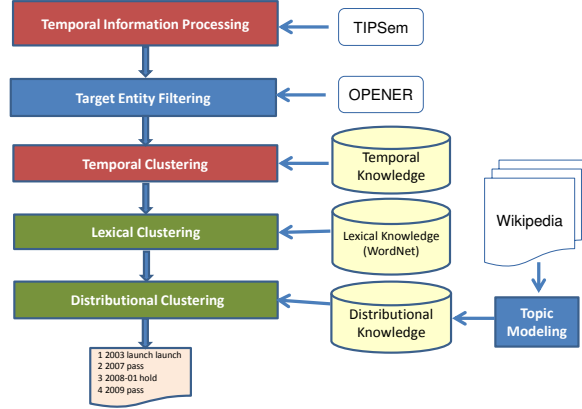


Figure 2: System architecture

3.1.1. Temporal Information Processing

Being the input of the system a set of plain texts, the events in those texts must be extracted. Furthermore, considering that the final aim is to build a timeline, temporal expressions and temporal links between events and times are required. For this reason, the first module of the proposed architecture is performing Temporal Information Processing. TIPSem system (Temporal Information Processing using Semantics) [37, 36]⁶ is used for this purpose. This system is based on morphosyntactic knowledge plus semantic knowledge, specifically, semantic networks and semantic roles. TIPSem is able to automatically annotate all the temporal information according to TimeML standard annotation scheme [54], which means annotating all the temporal expressions (TIMEX3), events (EVENT) and links between them.

3.1.2. Target Entity Filtering

Considering that not all the events annotated by the previous module are necessary to build the timeline, but only the ones related to a target entity, a Target Entity Filtering needs to be performed in order to avoid those events that are annotated but not related to the given entity.

The Target Entity Filtering requires resolving name entity recognition and entity coreference resolution. Since this is not the main challenge of our research, this task is performed using an external tool. Therefore, the

⁶<http://gplsi.dlsi.ua.es/demos/TIMEE/>

OPENER⁷ web services were integrated in our proposal. More specifically, the NER⁸ and the coreference resolution component⁹.

In order to determine if an event must be part of the timeline or not, this module chooses the events in which a target entity (or a target entity coreference) explicitly participates in a *has-participant* relation with the semantic role A0 (i.e. agent) or A1 (i.e. patient), as defined in the Propbank Project [49]. In case of nominal events, since the information of A0 or A1 is not obtained, this module chooses the events that have the target entity in the same sentence. Otherwise, the event is avoided.

3.1.3. Temporal Clustering Approach

As presented by Danlos (2003) [16], working on temporal relations between two events $e1$ and $e2$ determines if one event precedes, includes or overlaps the other one, supposing that $e1 \neq e2$ for all these temporal relations. However, for event coreference purposes we are interested in temporal relations that denote that $e1 = e2$. We named this as temporal compatibility¹⁰.

Considering these premises and using the temporal information knowledge extracted in the first module of the proposal, the temporal clustering algorithm is performed in two steps:

- *Within-document temporal clustering*: For each document, the temporal information of each event is going to be extracted. Each event is anchored to a time anchor¹¹ when a temporal SIMULTANEOUS/BEGIN/ INCLUDES link exists between this event and a temporal expression. After this, two events will be considered part of the same cluster if they are temporally compatible. This means that: a) two events are anchored to the same time anchor, or b) two events have a temporal SIMULTANEOUS link between them.

⁷<http://www.opener-project.eu/webservices>

⁸<http://opener.olery.com/ner>

⁹<http://opener.olery.com/coreference>

¹⁰SemEval2015 decided to simplify the representation of durative events in the timelines by anchoring them in time considering their starting point, so relation type BEGUNBY or INCLUDES have the same meaning as SIMULTANEOUS.

¹¹A time anchor is always a DATE (as defined in TimeML) and its format follows the ISO-8601 standard: YYYY-MM-DD, being the maximum granularity admitted in the task DAY. For lower granularity, only months and years are admitted.

Example 4 shows two events temporally compatible and clustered together.

- (4) a. The `<EVENT eid="e1"> meeting </EVENT>` was `<TIME3 tid="t1" value="2014-03-22"> yesterday </TIME3>`.
 b. At the same time, the teacher `<EVENT eid="e2"> presents </EVENT>` the ideas.
`<TLINK eventInstanceID="e1" relatedToTime="t1" relType="IS_INCLUDED" /> <TLINK eventInstanceID="e2" relatedToEventInstance="e1" relType="SIMULTANEOUS"/>`

Two events non-temporally compatible are shown in Example 5.

- (5) a. The `<EVENT eid="e1"> meeting </EVENT>` was `<TIME3 tid="t1" value="2014-03-22T17:00"> yesterday at 17:00 </TIME3>`.
 b. After that, the teacher `<EVENT eid="e2"> presents </EVENT>` the ideas.
`<TLINK eventInstanceID="e1" relatedToTime="t1" relType="IS_INCLUDED" /> <TLINK eventInstanceID="e2" relatedToEventInstance="e1" relType="AFTER"/>`

- *Cross-document temporal clustering*: From a set of documents (related by the same topic), and considering that in the previous step all the events were assigned a time anchor, all the events in the different documents that are temporally compatible will be clustered together.

- (6) a. Document 1: The `<EVENT eid="e1"> meeting </EVENT>` was `<TIME3 tid="t1" value="2014-03-22"> yesterday </TIME3>`.
`<TLINK eventInstanceID="e1" relatedToTime="t1" relType="IS_INCLUDED" />`
 b. Document 2: The students `<EVENT eid="e5"> met </EVENT>` on `<TIME3 tid="t3" value="2014-03-22">Tuesday</TIME3>`.
`<TLINK eventInstanceID="e5" relatedToTime="t3" relType="IS_INCLUDED" />`

According to Example 6, in the within-document temporal clustering, doc1-e1 is anchored to the date “2014-03-22”, and doc2-e5 is anchored to the same date. Therefore, in the cross-document temporal clustering step these two events will be considered part of the same cluster.

According to the guidelines for the task, in the case of different granularity, such as “2001” and “2001-05”, events with lower granularity should be given preference. However, as the task is defined, all the event clusters must be associated to only one time anchor. This fact poses a problem when we have two events that corefer and are anchored to compatible time anchors but different granularities: it is not possible to automatically cluster them together following the guidelines to build the Timelines. For example, having two sentences such as: “*Mary came to Spain in 2001*” and “*Mary visited*

Spain in May 2001”, it is impossible to automatically group these two events in the same cluster because there is a different granularity for the time anchors “2001” and “2001-05”. This problem will be studied and resolved in further proposals of the task.

3.1.4. Event Clustering Approach through lexical knowledge

Once all event mentions that refer to the same lapse of time have been grouped together, our system tries to detect those events that refer to the same facts (that is, those coreferent event mentions) using the semantic aspects of the event structure to its advantage. As we said before, we apply two types of semantic information: lexical and distributional. In this section, we introduce the event clustering based on lexical knowledge; in the next section, we will introduce the event clustering based on distributional knowledge.

The lexical approach tries to find coreferent events through the lexical information of the event head. The event head is the word or multiword expression that conveys the main aspect of the event. Example 7 shows again an event mention with the event head explicitly marked.

- (7) The Airbus A380, the world’s largest passenger plane, was set to <EVENT eid="e78">land</EVENT> in the United States of America on Monday after a test flight.

Our hypothesis in this lexical approach is that two event mentions corefer if, besides expressing the same time, their event head express the same concept. Specifically, two event heads will express the same concept if:

- both event heads are the same word (that is, they have the same lemma, see Example 8), and
- both event heads are synonyms (Example 9).

- (8) a. The airplane will be <EVENT eid="e89">carrying</EVENT> about 500 people.
 b. It is being billed as the first time it has <EVENT eid="e88">carried</EVENT> a near-normal number of passengers.
- (9) a. US automaker GM <EVENT eid="e88">reports</EVENT> losses of \$ 6 billion.
 b. United States automobile company General Motors <EVENT eid="e67">announced</EVENT> it has lost US\$ 6 billion in the first quarter of 2009, amidst heavy declines in revenues.
 c. The firm <EVENT eid="e75">said</EVENT> it had lost a net \$ 5.9 billion dollars, or \$ 9.66 per share.

Example 9 shows three sentences whose event heads are synonyms (*report*, *announce*, *say*) and express coreferent events. They are related to the same meaning of “saying”.

In order to extract lemmas and synonymy relations, the system uses WordNet [17] as a lexical knowledge base. This knowledge represents all the lexical knowledge that a human being needs during a language communication to interpret that two event mentions have the same fact as their reference. Therefore, the event heads of example 9 share the same synset in WordNet.

3.1.5. Event Clustering Approach through distributional knowledge

Sometimes, only semantic information about the event head is not enough to find coreferent events. Example 10 shows two sentences with the same temporal reference and the same lemma at the event head (“to state”), but which refer to different facts.

- (10) a. <A0>He</A0> <EVENT eid="e200">stated</EVENT> that <A1>his intent was not to “perpetuate the bad business decisions of the past” </A1>.
- b. Of the government ownership <A0>he</A0> <EVENT eid="e179">stated</EVENT> that <A1>he refused “to let General Motors and Chrysler become wards of the state” </A1>.

In order to distinguish between these non coreferential events and the coreferential ones, it is necessary to also consider the semantics of the arguments of the event structure. The arguments of the event structure consist of those elements that complete the event (called semantic roles from a linguistic point of view). In example 10, the event “to state”, in order to be a complete event, always needs at least “a person who states” (argument calls agent or A0) and “something stated” (argument calls theme or A1). In this example, both events are not coreferent because argument A1 is semantically different.

Using the distributional approach, it is also possible to find coreferential events whose event heads are not synonyms or the same lemma (Example 11):

- (11) a. Bank of America <EVENT eid="e86">reports</EVENT> losses of over US\$ 2.2 billion.
- b. The Bank of America Corporation has <EVENT eid="e82">announced</EVENT> that it lost US\$ 2.24 billion in the third quarter of this year.

In this example, the event heads “report” and “announce” are not synonyms (they do not share the same synset in WordNet). However, they are

coreferent because, according to their arguments, they are referring to the same fact, and the distributional clustering is able to detect it.

It is not clear exactly which components of event structure are the determining factor in event coreferent resolution [13]. Rather than creating a complex feature matrix to represent the semantic of the argument as [6] does, we propose a compact, used-based distributional representation of the semantics of the arguments. Moreover, contrary to [22], who uses a syntax-based distributional representation, we apply a semantic roles-based distributional approach.

Assuming that Wikipedia represents all the previous experiences of a speaker, our system uses a topic-based distributional semantic representation which applies LDA Topic Modeling [8] to Wikipedia. LDA Topic Modeling extracts a set of topics from large corpora. Each topic is a distribution over a fixed vocabulary, so that each one is represented by the most frequent words. Therefore, applying LDA over Wikipedia, we obtain a lemma-topic matrix in which the distributional meaning of each word is represented by a vector of topics: the topics in which a lemma tends to appear. This lemma-topic matrix is used as a distributional knowledge base. All the experiments have been done with 500 topics.

Formally, the distributional knowledge base is a matrix $D \in M_{n \times m}(\mathbb{R})$ where n stands for the most frequent lemmas on Wikipedia and m is the amount of topics extracted (500). Each row D_n , that represents a lemma, is a vector formed by the weight of the lemma in each topic. Table 1 shows a sample of three possible topic models extracted from Wikipedia.

Table 1: Topic Models from Wikipedia

Topic	Words
0	smiling blood solidity diuretic rinne thermophiles...
1	apple macbook battery pro keynote notebook features...
2	plane aircraft crash flight air airport passengers accident ...

Once the distributional knowledge base is created, the coreference resolution process is as follows: we extract the event structure for each sentence (the event head and the arguments). For each content word of the arguments (nouns, adjectives, etc.), the system extracts its distributional vector from the distributional knowledge base. In this way we obtain a vector of topics for the words of each argument.

Formally, the event structure is represented as a tuple of three elements: the argument $A0$, the argument $A1$ and the event head H :

$$ES = \langle A0, A1, H \rangle$$

An event argument is a set of content words. For each word w_n , the system extracts its distributional topic-based vector from the Wikipedia distributional knowledge base ($\vec{V}(w_n)$). An argument is then a set of vectors corresponding to each word in the argument set:

$$A0 = \{\vec{V}(w_1), \dots, \vec{V}(w_n)\}$$

Thereupon the distributional meaning of all the arguments of an event is obtained by compositionality. In this case, we have applied a simple additive method [45] to obtain a single vector that represents the meaning of the whole event structure. Let $\vec{V}(A0)$ be the compositional vector of an argument $A0$, it is calculated as:

$$\vec{V}(A0) = \sum^n \vec{V}(w_n)$$

Finally, the compositional vector of the whole event structure $\vec{V}(ES)$ is:

$$\vec{V}(ES) = \vec{V}(A0) + \vec{V}(A1)$$

With this vector, the distributional semantic representation of the event structure is finished. It represents all the topics related to the words of the arguments in a specific event mention. Thus we have all the distributional information of the event structure contained in a single vector.

In order to determine if two events are coreferential, the system calculates the cosine similarity between both event vectors. If the cosine similarity between two event vectors is higher than 0.9¹², the system concludes that there is a coreference between them and hence they are grouped together in the same cluster. Formally

$$coref(\vec{V}(ES_1), \vec{V}(ES_2)) \implies sim(\vec{V}(ES_1), \vec{V}(ES_2)) \geq 0.9$$

We consider coreference as a transitive relation, therefore:

$$coref(\vec{V}(ES_a), \vec{V}(ES_b)) \wedge coref(\vec{V}(ES_b), \vec{V}(ES_c)) \implies coref(\vec{V}(ES_a), \vec{V}(ES_c))$$

¹²After some tests, we have settle a threshold of 0.9 over 1.

4. Evaluation

As explained before, given a set of documents and a set of target entities, the task of event ordering consists in building a timeline related to each entity, i.e. detecting, anchoring in time, and ordering the events in which the target entity is involved [43]. This research is focused on written news in English, since news describe actions and the relationships between them. However, it could be extended to any other domains.

4.1. Evaluation Environment

In order to evaluate our approach, the dataset provided for Task4 at SemEval 2015 has been used.¹³ The dataset used for this task was composed of articles from Wikinews about three topics: a) Airbus and Boeing (corpus 1); b) General Motors, Chrysler and Ford (corpus 2); and c) Stock Market (corpus 3).

These evaluation corpora consist of 30 documents for each corpus (around 30,000 tokens and 915 events altogether) and they are very similar in terms of size. It is interesting to notice, however, that the timelines created from Stock Market corpora have a lower average number of events (20.3 events) with regard to those created from the other corpora (26.4 events for Airbus and 25.7 events for GM). It is important also to emphasize that, on average, Stock Market timelines contain events from a higher number of different documents, i.e. 9.1 versus 6.2 for Airbus and 5.7 for GM.

At SemEval 2015 Task 4, two different tracks were proposed on the basis of the data used as input: **Track A** for which they provided only raw text sources, and **Track B**, for which they also provided available gold events mentions. All the experiments shown in this section were performed using Track B input.

The evaluation metrics used at SemEval 2015-Task 4 are based on the evaluation metrics used for TempEval-3. They defined temporal awareness [60] as the performance that implies the correct recognition and classification of the temporal entities involved in the temporal relations. In order to perform the evaluation of the temporal awareness, each timeline is transformed by the evaluation tool into a set of temporal relations [43]. Then, precision, recall and F1-score are calculated.

¹³<http://alt.qcri.org/semeval2015/task4/index.php?id=data>

Furthermore, due to the fact that we are evaluating three different corpora, and different timelines for each corpora, in which the number of events involved is different, the results obtained in the evaluation are presented in terms of micro-average, due to the fact that this measure is useful when the dataset varies in size. In micro-average, we are not considering just an average of the precision and recall of the different sets, but rather we are summing up the individual true positives, false positives, and false negatives of the system for the different sets and applying them to obtain precision, recall and F1 measures.

4.2. Experiments and Results

For our evaluation, four different experiments have been performed in order to determine what is the best way to deal with event ordering task. The descriptions of the experiments are:

- TC+LCV1: Temporal clustering + Lexical Semantic clustering considering only lemmas
- TC+LCV2: Temporal clustering + Lexical Semantic clustering considering lemmas and synonyms
- TC+DSC: Temporal clustering + Distributional Semantic clustering
- TC+LCV2+DSC: Temporal clustering + Lexical Semantic clustering considering lemmas and synonyms+ Distributional Semantic Clustering

The Track B¹⁴ results obtained by the system in the evaluation of Micro-F1, Micro-Precision and Micro-Recall are presented in Table 2 per subcorpora and over the whole dataset. The experiment TC+LCV1 is the same as the one called Run1 in our participation at SemEval2015-Task 4. The other three experiments are the novelty of this work. Furthermore, the comparison between our results and the results obtained by other systems can only be done with the HEIDELTOUL team's outcome, since it was the only team that participated in the same Track as our system at SemEval 2015-Task 4. Due to the fact that this is a very novel task, no other competition has been held. The HEIDELTOUL team presented also two different runs, whose results are also presented in this table.

¹⁴Timelines with time anchors from texts annotated with events.

Measure	Approach	Airbus	GM	Stock	Total
<i>Micro-F1</i>	TC+LCV1	22.35	19.28	33.59	25.36
	TC+LCV2	26.01	22.13	31.39	26.53
	TC+DSC	25.19	19.62	31.71	26.20
	TC+LCV2+DSC	26.21	21.08	31.58	26.61
	HEIDELTOUL1	19.62	7.25	20.37	17.03
	HEIDELTOUL2	16.50	10.94	25.89	18.34
<i>Micro-P</i>	TC+LCV1	17.73	14.25	36.58	21.73
	TC+LCV2	20.58	16.60	36.50	23.56
	TC+DSC	19.85	14.12	37.81	23.29
	TC+LCV2+DSC	20.80	15.69	37.04	23.68
	HEIDELTOUL1	17.75	9.46	34.02	20.11
	HEIDELTOUL2	10.82	8.84	21.97	13.58
<i>Micro-R</i>	TC+LCV1	30.20	29.78	31.06	30.46
	TC+LCV2	35.33	33.21	27.53	30.37
	TC+DSC	34.47	32.12	27.31	29.96
	TC+LCV2+DSC	35.43	32.12	27.53	30.37
	HEIDELTOUL1	21.94	5.88	14.54	14.76
	HEIDELTOUL2	34.76	14.34	31.50	28.23

Table 2: Results on the SemEval-2015 Task4 TrackB

The four experiments have been evaluated more in detail (Table 3) considering the selection of events in which a target entity is involved, regardless of their ordering in timelines. The number of true positives and F1-scores are shown. Besides, the evaluation of the time anchors assignment in terms of accuracy is also provided¹⁵.

Runs	Airbus			GM			Stock			Total		
	Events		TA	Events		TA	Events		TA	Events		TA
	TP	F1	Acc	TP	F1	Acc	TP	F1	Acc	TP	F1	Acc
<i>TC+LCV1</i>	240	59.33	36.67	234	67.73	24.34	190	72.80	43.16	664	65.68	34.17
<i>TC+LCV2</i>	247	60.30	40.00	244	68.93	27.00	181	69.08	43.00	672	65.56	36.00
<i>TC+DSC</i>	241	59.06	39.83	243	68.64	27.16	180	68.70	43.33	664	64.84	36.14
<i>TC+LCV2+DSC</i>	247	60.39	39.67	244	69.02	27.04	181	69.08	43.09	672	65.59	36.01

Table 3: Evaluation of the selection of events in which a target entity is involved and of time anchors assignment

Regarding other systems presented in the background that are performing event coreference, the comparison is not possible since these systems are only resolving event coreference but not event ordering, and using different corpora

¹⁵For each timeline, the accuracy is computed by dividing the number of matching events/time anchors by the number of correctly identified events (TP in table).

in order to evaluate event coreference. [51] is using Ontonotes corpora, [12] and [42] are using ACE corpus, [5] used also ACE corpora and ECB corpora, [32] is using ECB corpora, [35] is using Intelligence Community (IC) corpora, [14] and [22] are using both IC and ECB corpora. [34] and [4] corpora are unavailable.

In the next section, we will analyze the results obtained from the different experiments performed and we will compare these results with the current state-of-the-art systems.

4.3. Discussion

Table 2 presents the results of evaluating our experiments in the cross-document event ordering task framework provided by SemEval2015 Task 4 (Track B). The best performance is achieved by the combination of the temporal clustering, the lexical semantic clustering using lemmas and synonyms and the distributional semantic clustering (26.61% in F1-score).

F1-score results are quite similar between TC+LCV2 and TC+LCV2+DSC experiments. It is remarkable that the three-cluster experiment is obtaining the best precision results. It should be pointed out too that all our experiments outperform the state-of-the-art (system HEIDELTOUL) in all its runs and all metrics, with a difference of 8.27 points in the F1-score when we compare the best solutions by both systems.

Although the results obtained by the temporal plus distributional semantic clustering are reasonable by themselves, it is remarkable that there is an increase both in precision and recall when adding the lexical clustering to this experiment instead of performing TC+DSC alone.

The system is performing better on the “Stock Market” corpus. One of the reasons is that in the timelines related to this corpus all the events were ordered, while in the other two corpora less than 70% of the events were ordered. Since our approach is using the TipSEM system to determine time anchors and time relations, most of the events in the timelines are anchored to a date and therefore they are ordered in time. However, in the gold standard timelines for “Airbus” and “GM”, 30% of the events in the timeline are in position 0 with an undefined date (XXXX-XX-XX) and in these cases, if the system detects the event, it is not considered a true positive.

Despite the fact that adding distributional semantic clustering to the temporal and lexical clustering approach improves the precision of the system, the improvement is less than we expected. After these experiments we conclude that the distributional representation based on Topic Models is

too abstract for event coreference resolution. In the end, many words are related to the same topic model. This is why the similarity threshold has been fixed in 0.9: a low threshold will group many non-coreferential events together. Therefore, although distributional representation based on Topic Models is right (there is a clear improvement compared to other approaches and to the lemma-based approach), it suffers from over-representation. The challenge for Future Works is to find a vectorial semantic representation that is more specific than the topic-modeling approach, but more abstract than the lemma- and synonym-based approach.

The figures in Table 3 show an F1-score higher than 64% for all the corpora in the case of selected events involving a target entity in all the experiments. In this evaluation, the ordering is not being considered and this is why F1-score is higher than in Table 2 where all the elements related to event ordering are taken into account.

Regarding the accuracy results of time anchors, all our experiments are obtaining very similar results (up to 34%). This is consistent since the temporal clustering, performed by the TipSEM system, is the same for all of the experiments and this accuracy is totally dependent on the performance of the TipSEM tool in determining proper temporal relations and normalization of temporal expressions.

According to Tempeval-3 [60] evaluation, the TipSEM system is obtaining an F1-score of 65.31% at temporal expression performance and an F1-score of 42.39% at temporal awareness regarding temporal relations. In event ordering both tasks are combined with a final F1-score of 36% in the whole, for all the experiments except from the one using only temporal and lemmas clustering that is obtaining 34.17%. The improvement in accuracy for the three new experiments is due to a granularity problem of the experiment presented in SemEval2015-Task 4 (same as TC+LCV1). The problem was due to the fact that TipSEM normalization was resolving temporal expressions according to the granularity of the expression itself, i.e. “three months ago” as “YYYY-MM”. However, all the expressions in the output of the SemEval2015-Task 4 were transformed to the granularity format “YYYY-MM-DD”. This was adapted for the new experiments causing thus an increase in the accuracy measure.

After analyzing the main problems of the different experiments performed, three main types of errors were detected:

- Errors due to a wrong temporal expression normalization or a wrong

temporal relation assignment obtained by the TipSEM system. Since the temporal clustering is the first step of the system, only those events with temporal compatibility (events happening at the same moment) are clustered together. However, if the TipSem system wrongly detects the time anchor of an event, it will cluster it wrongly, provoking a chain of errors at temporal anchoring, event coreference and ordering in the timeline. All the experiments will suffer from this type of error since the temporal clustering is the first part of all the experiments in our system.

- Errors due to a wrong detection of the arguments of the event. When adding distributional semantic clustering to the lexical clustering, if two events are considered coreferent by the lexical clustering but some of the arguments of one of them (A0 or A1) are pronouns referring to the target entity, the topics in the vector are, most of the times, not properly generated and therefore those events are not considered coreferent by the distributional semantic clustering and they will be separated in two different clusters.

Example (12) shows two coreferent events that share the same event head (“to offer”). The lexical cluster groups them together correctly as coreferential. However, due to the fact that, on one hand, the role A0 of the second event is an anaphora (“both”, which refers to Chrysler and GM) and, on the other hand, practically no words of roles A1 are semantically related (maybe only “auto” and “vehicle”), the distributional cluster finally splits them incorrectly into two different clusters. The distributional cluster has not enough distributional semantic information to detect the event coreference. An anaphora resolution system will resolve, partially, this problem.

- (12) a. <A0>Chrysler</A0> will <EVENT eid="e88">offer</EVENT> <A1>a \$25,000 vehicle voucher...</A1>
 b. <A0>Both</A0> will <EVENT eid="e96">offer</EVENT> <A1>the deal to most United Auto Workers (UAW) union members</A1>.

This entity coreference resolution problem is also impoverishing the results in the evaluation of the event selection in which the target entity is involved as shown in Table 3. This is an important source of errors in the selection of events since there are a lot of anaphoric references in texts from newspapers.

In order to analyse in depth the impact of anaphoric expressions in our experiments, we have extracted the manual annotation of the evaluation corpora provided by the MEANTIME corpus[44]. The MEANTIME Corpus (the NewsReader Multilingual Event AND TIME Corpus) consists of the same corpora as the one used at SemEval2015-Task 4 and their translations in Spanish, Italian, and Dutch. It has been annotated manually at multiple levels, including entities, events, temporal information, semantic roles and entity coreference. For our analysis, we have extracted this entity coreference annotation, used it for both the distributional semantic clustering and selection of events, and re-evaluated our three main approaches using them, in order to determine if there is an improvement. For this experiment, only the anaphoric expressions related to the entities of the timelines have been considered and the results are presented in Table 4.

Approach	Micro-F1	Micro-P	Micro-R
TC+LCV2+MeantimeEC	29.95	25.17	36.98
TC+DSC+MeantimeEC	29.73	24.94	36.78
TC+LCV2+DSC+MeantimeEC	30.09	25.37	36.98

Table 4: Results using MEANTIME manual entity coreference annotation

As shown in Table 4, the best experiment is still the one combining temporal, lexical and distributional semantics but all of the approaches obtain better results in all measures, and especially in terms of recall, since there are many more properly selected events when the entity coreference resolution is applied. Our best approach (TC+LCV2+DSC) is obtaining an increase of 3.5 points in F1 measure, 1.7 in precision and the most outstanding value, 6.61 points in recall.

- Similar problems arise with regard to nominal events: event structures whose event head is not a verb but a noun, as the event b. in Example (13)

- (13) a. <A0>the first Airbus A380</A0> <EVENT eid="e93">landed</EVENT> <A1>at Singapore's Changi International Airport</A1> at 6:40 p.m. (GMT+8).
 b. the <EVENT eid="e74">landing</EVENT> was a milestone not only for SIA, but also for Changi International Airport.

The problem is that, due to the fact that the event “the landing” is a noun (a deverbal noun to be precise, a noun derived from a verb),

it is not possible to extract semantic roles from it. The distributional representation of these events is formed only by the topic vector of the noun head. Therefore, their semantic representativity is smaller, compared to the compositional vector of verbal events. This is why nominal events have not enough semantic information for their coreferences to be detected. Nominal events coreference needs to be detected by means of a specific strategy.

This problem also affects the results in the evaluation of the selection of events, because according to the guidelines for this task, only the events in which the target entity explicitly participates in a *has participant* relation, with the semantic role A0 or A1, must be selected. In case of nominal events, our approach consists in selecting the event if the target entity is contained in the sentence but this is not always true.

In order to analyse more deeply the impact of nominalization in our experiments results, we have extracted the number of nominal events from the manual annotation of event mentions at MEANTIME corpus [44]. In this corpus, the event mentions have an attribute Part of Speech (POS) with “NOUN” value for nominal events. From a total of 992 events in the Gold Standard output, 132 of them were considered as nominal events according to MEANTIME manual annotation. Therefore, around a 13% of the total events in the output were nominal events¹⁶.

The results reveal that the combination of a temporal clustering, a lexical clustering including lemmas and synonyms and a distributional clustering, consistently outperforms the other experiments. It must be also taken into account that the combination of the distributional semantic clustering plus the lexical semantic clustering is improving the results for all measures compared with the distributional semantic clustering in isolation, and all of the experiments presented here are also improving the results of the state-of-the-art systems. Despite the fact that distributional clustering suffers from errors of external tools (anaphora resolution, semantic roles labeling) and from under-representation (nominal events), improving or solving these problems

¹⁶Since the POS attribute is not compulsory in MEANTIME annotation guidelines, some event mentions have no POS attribute, meaning that probably there are many more nominal events than the ones that have been annotated

will show that this combination of different semantic clustering benefits, even more, the performance of the event ordering task resolution over all metrics rather consistently.

5. Conclusions

In this work we present a system that is able to cope with the cross-document event ordering task. This task involves dealing with automatic event extraction, cross-document temporal relation extraction and cross-document event coreference resolution. In order to tackle these problems, three different types of knowledge have been used:

- Temporal knowledge that allows the system to perform a temporal clustering based on the temporal compatibility of two events, implying that two events are part of the same cluster if they happen at the same time. This inference is being performed both at within- and cross-document level.
- Lexical-semantic knowledge that allows the system to determine whether two events refer to the same fact if their event heads express the same concept. We are assuming that this happens when their event heads either have the same lemma or are synonyms.
- Distributional semantic knowledge that considers the semantics of the arguments of the event structure, allowing the system to determine if two events are coreferent because they are semantically compatible and usually appear in the same context.

In order to analyse the impact of these different types of knowledge in the event ordering task, the system has been evaluated under the framework proposed at SemEval-2015 Task 4. This task consists in generating timelines of events related to a target entity. Four different experiments have been performed. All of them include the temporal clustering since it is impossible for two events to be coreferent if they occur at different times. Then, we have combined: a) the temporal clustering with the lexical-semantic clustering (using lemmas and synonyms), b) the temporal clustering with the distributional-semantic clustering, and c) the temporal clustering with the lexical-semantic clustering and the distributional semantic clustering.

Results show that the timeline creation is a very challenging task (F1-Score of 26.61%) but with our approach we are outperforming the results

of the state-of-the-art systems (+8.27 points in F1-score) and we consider that the combination of temporal knowledge, lexical semantic knowledge and distributional semantic knowledge achieves reasonable results even when the distributional semantic clustering is very dependent on external tools such as entity coreference resolution and semantic role labeling. Besides, nominal events require a special treatment, which is beyond the scope of this paper. Therefore, we can conclude that the combination of temporal, lexical-semantic and distributional information is a proper approach to Cross-Document Event Ordering since it is outperforming the current state-of-the-art systems.

Obviously, timeline creation is an open task and in future research we would like to explore some other strategies for event ordering tasks, such as combining the clustering process in a different way, dealing with nominal events, improving entity coreference resolution to properly represent the arguments of the events and improving distributional representation through word embeddings. Furthermore, it is also our goal to tune the system in order to evaluate it within the Track A framework of SemEval2015 Task 4, and compare the results with the teams that participated in said Track.

References

- [1] Ahn, D., 2006. The stages of event extraction. In: Proceedings of the Workshop on Annotating and Reasoning About Time and Events. ARTE '06. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 1–8.
URL <http://dl.acm.org/citation.cfm?id=1629235.1629236>
- [2] Alonso, O., Gertz, M., Baeza-Yates, R., Dec. 2007. On the value of temporal information in information retrieval. SIGIR Forum 41 (2), 35–41.
- [3] Bach, E., 1986. The Algebra of Events. Linguistics and Philosophy 9, 5–16.
- [4] Bagga, A., Baldwin, B., 1999. Cross-document event coreference: Annotations, experiments, and observations. In: In Proc. ACL-99 Workshop on Coreference and Its Applications. pp. 1–8.
- [5] Bejan, C., Harabagiu, S., July 2010. Unsupervised event coreference resolution with rich linguistic features. In: Proceedings of the 48th Annual

Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Uppsala, Sweden, pp. 1412–1422.

URL <http://www.aclweb.org/anthology/P10-1143>

- [6] Bejan, C. A., Harabagiu, S., 2014. Unsupervised Event Coreference Resolution. *Computational Linguistics* 40 (2), 311–347.
- [7] Bethard, S., 2013. Cleartk-timeml: A minimalist approach to tempeval 2013. In: *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Association for Computational Linguistics, Atlanta, Georgia, USA, pp. 10–14.
- [8] Blei, D. M., Ng, A. Y., Jordan, M. I., 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, 993–1022.
- [9] Caselli, T., Fokkens, A., Morante, R., Vossen, P., June 2015. Spinoza_vu: An nlp pipeline for cross document timelines. In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Association for Computational Linguistics, Denver, Colorado, pp. 787–791.
- [10] Chambers, N., June 2013. Navytime: Event and time ordering from raw text. In: *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Association for Computational Linguistics, Atlanta, Georgia, USA, pp. 73–77.
- [11] Chang, A., Manning, C. D., June 2013. SUTime: Evaluation in tempeval-3. In: *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Association for Computational Linguistics, Atlanta, Georgia, USA, pp. 78–82.
- [12] Chen, Z., Ji, H., 2009. Graph-based event coreference resolution. In: *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing. TextGraphs-4*. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 54–57.
URL <http://dl.acm.org/citation.cfm?id=1708124.1708135>
- [13] Cybulska, A., Vossen, P., 2013. Semantic relations between events and their time, locations and participants for event coreference resolution. In: *RANLP. RANLP 2011 Organising Committee / ACL*, pp. 156–163.

- [14] Cybulska, A., Vossen, P., May 2014. Using a sledgehammer to crack a nut? lexical diversity and event coreference resolution. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). European Language Resources Association (ELRA), Reykjavik, Iceland, pp. 4545–4552, aCL Anthology Identifier: L14-1646.
- [15] Daniel, N., Radev, D., Allison, T., 2003. Sub-event based multi-document summarization. In: HLT-NAACL Text summarization workshop. ACL, pp. 9–16.
- [16] Danlos, L., 2003. Event Coreference between Two Sentences. *Computing Meaning* 77, 271–288.
- [17] Fellbaum, C., 1998. WordNet: An Electronic Lexical Database (Language, Speech, and Communication). MIT Press.
- [18] Ferro, L., Gerber, L., Mani, I., Sundheim, B., Wilson, G., 2005. TIDES 2005 Standard for the Annotation of Temporal Expressions. Tech. rep., MITRE.
- [19] Filannino, M., Brown, G., Nenadic, G., 2013. Mantine: Temporal expression identification and normalization in the tempeval-3 challenge. In: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). Association for Computational Linguistics, pp. 53–57.
- [20] Firth, J. R., 1957. *Papers in Linguistics (1934-1951)*. Oxford University Press, Oxford.
- [21] Gärdenfors, P., 2014. *The geometry of meaning. Semantics based on conceptual spaces*. MIT Press, Cambridge, Mass.
- [22] Goyal, K., Jauhar, S. K., Li, H., Sachan, M., Srivastava, S., Hovy, E. H., 2013. A structured distributional semantic model for event co-reference. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 2: Short Papers. The Association for Computer Linguistics, pp. 467–473.
URL <http://aclweb.org/anthology/P/P13/>

- [23] Grishman, R., Sundheim, B., 1996. Message understanding conference-6: A brief history. In: COLING. pp. 466–471.
- [24] Group, I. T. W., 2008. ISO TimeML TC37 draft international standard DIS 24617-1.
URL <http://semantic-annotation.uvt.nl/ISO-TimeML-08-13-2008-vankiyong.pdf>.
- [25] Harris, Z., 1968. Mathematical structures of language. Wiley, New York.
- [26] Hovav, M. R., Doron, E., Sichel, I., 2010. Lexical Semantics, Syntax, and Event Structure. Oxford University Press, Oxford.
- [27] Humphreys, K., Gaizauskas, R., Azzam, S., 1997. Event coreference for information extraction. In: Proceedings of the Workshop on operational factors in practical, robust anaphora resolution for unrestricted texts, 35TH meeting of ACL. pp. 75–81.
- [28] Ji, H., Grishman, R., Chen, Z., Gupta, P., 2009. Cross-document event extraction and tracking: Task, evaluation, techniques and challenges. In: RANLP. RANLP 2009 Organising Committee / ACL, pp. 166–172.
- [29] Jung, H., Stent, A., June 2013. Att1: Temporal annotation using big windows and rich syntactic and semantic features. In: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). Association for Computational Linguistics, Atlanta, Georgia, USA, pp. 20–24.
- [30] Kolomiyets, O., Moens, M.-F., June 2013. Kul: Data-driven approach to temporal parsing of newswire articles. In: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). Association for Computational Linguistics, Atlanta, Georgia, USA, pp. 83–87.
- [31] Laparra, E., Aldabe, I., Rigau, G., 2015. Document level time-anchoring for timeline extraction. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015). Beijing, China.

- [32] Lee, H., Recasens, M., Chang, A., Surdeanu, M., Jurafsky, D., 2012. Joint entity and event coreference resolution across documents. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. EMNLP-CoNLL '12. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 489–500.
- [33] Levin, B., Rappaport-Hovav, M., 2005. Argument realization. Cambridge University Press, Cambridge (UK).
- [34] Li, P., Zhu, Q., Zhu, X., July 2011. A clustering and ranking based approach for multi-document event fusion. In: Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2011 12th ACIS International Conference on. pp. 159–165.
- [35] Liu, Z., Araki, J., Hovy, E. H., Mitamura, T., 2014. Supervised within-document event coreference using information propagation. In: Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (Eds.), Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Iceland, May 26-31, 2014. European Language Resources Association (ELRA), pp. 4539–4544.
URL <http://www.lrec-conf.org/lrec2014>
- [36] Llorens, H., Saquete, E., Navarro-Colorado, B., 2012. Automatic System for Identifying and Categorizing Temporal Relations in Natural Language. *International Journal of Intelligent Systems* 27 (7), 680–703.
- [37] Llorens, H., Saquete, E., Navarro-Colorado, B., 2013. Applying Semantic Knowledge to the Automatic Processing of Temporal Expressions and Events in Natural Language. *Information Processing & Management* 49 (1), 179–197.
- [38] Mani, I., Pustejovsky, J., Gaizauskas, R., 2005. The Language of Time. Oxford University Press, Oxford.
- [39] Mann, G. S., 2007. Multi-document relationship fusion via constraints on probabilistic databases. In: Sidner, C. L., Schultz, T., Stone, M., Zhai, C. (Eds.), Human Language Technology Conference of the North

- American Chapter of the Association of Computational Linguistics, Proceedings, April 22-27, 2007, Rochester, New York, USA. The Association for Computational Linguistics, pp. 332–339.
- [40] Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., McClosky, D., 2014. The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. pp. 55–60.
 - [41] March, O., Baldwin, T., 2008. Automatic event reference identification. In: ALTA 2008. Australia, pp. 79–87.
 - [42] McConky, K., Nagi, R., Sudit, M., Hughes, W., March 2012. Improving event co-reference by context extraction and dynamic feature weighting. In: Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2012 IEEE International Multi-Disciplinary Conference on. pp. 38–43.
 - [43] Minard, A.-L., Speranza, M., Agirre, E., and Marieke van Erp, I. A., Magnini, B., Rigau, G., Urizar, R., 2015. Semeval-2015 task 4: Timeline: Cross-document event ordering. In: Proceedings of the 9th International Workshop on Semantic Evaluation. SemEval '15. Association for Computational Linguistics, pp. 778–786.
 - [44] Minard, A.-L., Speranza, M., Urizar, R., Altuna, B., van Erp, M., Schoen, A., van Son, C., May 2016. Meantime, the newsreader multilingual event and time corpus. In: Proceedings of 10th edition of the Language Resources and Evaluation Conference (LREC). Slovenia.
 - [45] Mitchell, J., Lapata, M., 2010. Composition in Distributional Models of Semantics. *Cognitive Science* 34, 1388–1429.
 - [46] Moulahi, B., Strötgen, J., Gertz, M., Tamine, L., June 2015. Heidehtoul: A baseline approach for cross-document event ordering. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). Association for Computational Linguistics, Denver, Colorado, pp. 825–829.
 - [47] Navarro, B., Saquete, E., June 2015. Gplsiua: Combining temporal information and topic modeling for cross-document event ordering. In:

- Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). Association for Computational Linguistics, Denver, Colorado, pp. 820–824.
- [48] Navarro-Colorado, B., 2015. Towards Vector Space Models of Semantics and the Semiotic Textology. In: Borreguero, M., Vitacolonna, L. (Eds.), International Workshop Janos S. Petöfi, In Memoriam.
 - [49] Palmer, M., Gildea, D., Kingsbury, P., 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics* 31.
 - [50] Petöfi, J., 1991. Towards a Semiotic Theory of the Human Communication. Gold Press, Szeged.
 - [51] Pradhan, S. S., Ramshaw, L. A., Weischedel, R. M., MacBride, J., Micciulla, L., 2007. Unrestricted coreference: Identifying entities and events in ontonotes. In: ICSC. IEEE Computer Society, pp. 446–453.
 - [52] Pustejovsky, J., Castaño, J. M., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., Katz, G., 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. In: IWCS-5.
 - [53] Saquete, E., González, J. L. V., Martínez-Barco, P., Muñoz, R., Llorens, H., 2009. Enhancing QA Systems with Complex Temporal Question Processing Capabilities. *Journal of Artificial Intelligence Research (JAIR)* 35, 775–811.
 - [54] Saurí, R., Littman, J., Knippen, R., Gaizauskas, R., Setzer, A., Pustejovsky, J., 2006. TimeML Annotation Guidelines 1.2.1 (<http://www.timeml.org/>).
 - [55] Semeval-2015, 2015. International Workshop on Semantic Evaluation. URL <http://alt.qcri.org/semeval2015/>
 - [56] Setzer, A., 2001. Temporal information in newswire articles: An annotation scheme and corpus study. Ph.D. thesis, University of Sheffield.
 - [57] Strötgen, J., Zell, J., Gertz, M., 2013. Heideltime: Tuning english and developing spanish resources for tempeval-3. In: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). Association for Computational Linguistics, Atlanta, Georgia, USA, pp. 15–19.

- [58] Turney, P., Pantel, P., 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37, 141–188.
- [59] UzZaman, N., Allen, J. F., 2010. Trips and trios system for tempeval-2: Extracting temporal information from text. In: *Proceedings of the 5th International Workshop on Semantic Evaluation. SemEval '10*. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 276–283.
- [60] UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M., Pustejovsky, J., 2013. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In: *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. ACL, Atlanta, Georgia, USA, pp. 1–9, ISBN: 978-1-937284-49-7.
- [61] Verhagen, M., Gaizauskas, R., Hepple, M., Schilder, F., Katz, G., Pustejovsky, J., 2007. Semeval-2007 task 15: Tempeval temporal relation identification. In: *Proceedings of the 4th International Workshop on Semantic Evaluations*. ACL, Prague, pp. 75–80.
- [62] Verhagen, M., Saurí, R., Caselli, T., Pustejovsky, J., 2010. Semeval-2010 task 13: Tempeval-2. In: *Proceedings of the 5th International Workshop on Semantic Evaluation*. ACL, Uppsala, Sweden, pp. 57–62.
- [63] Zavarella, V., Tanev, H., June 2013. Fss-timex for tempeval-3: Extracting temporal information from text. In: *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Association for Computational Linguistics, Atlanta, Georgia, USA, pp. 58–63.